

河南大学 2024-2025 学年第一学期

《统计学》结课论文

如果感到幸福，你拍拍手

——基于中国综合社会调查的幸福感研究

论文作者姓名：买伊然 戚路逍 王潇艺

作 者 学 号：2228200325

2228200260

2228200338

所 在 学 院：欧亚国际学院

所 学 专 业：金融数学

教 师 姓 名：辛 欣

论 文 得 分：

2024 年 12 月 20 日

摘 要

国民幸福值，作为衡量一个国家民众福祉状况的核心指标，其重要性不言而喻。它不仅关乎民众的基本生活需求是否得到满足，更体现了民众在心理、社会、经济及环境等多个层面上的综合福祉状态。本研究致力于构建一个全面、多维的国民幸福感评估体系，以期能够更为精准地反映民众的真实福祉水平。

在构建评估体系的过程中，我们充分考虑了经济、社会、环境及心理等多个领域的各项指标与因素。通过系统性地整合这些复杂而多样的信息，我们成功构建了一个既全面又具可操作性的国民幸福感评估框架。这一框架不仅涵盖了民众在物质层面的生活状况，还深入探讨了其在精神、社交及自然环境等方面的福祉感受。

基于这一评估体系，我们进一步建立了线性回归模型，旨在从多个角度深入剖析影响国民幸福满意度的各种因素。通过这一模型，我们能够更加准确地识别出哪些因素对幸福值具有显著影响，以及这些影响是如何在不同情境下发挥作用的。

在研究过程中，我们得出了一系列重要的观察结果。首先，不同变量对幸福值的影响程度存在显著的差异。这意味着，在制定提升国民福祉的政策措施时，需要充分考虑各变量的独特作用，以实现精准施策。其次，深入理解各变量的具体含义，对于准确把握幸福值的构成及变动规律至关重要。这有助于我们更全面地了解民众福祉的实际情况，为制定更加科学合理的政策提供有力支持。

此外，为了直观展现居民对生活的幸福感知，我们设定了五个幸福值指标。这些指标涵盖了民众在多个方面的福祉感受，能够更全面地反映其生活质量的真实状况。同时，通过分析八个变量间的相关性矩阵，我们获得了关于这些变量相互关系的宝贵信息。这些信息对于揭示幸福值背后的复杂机制、指导后续的实验研究具有重要的参考价值。

最后，我们利用箱线图和堆积频数分布直方图等统计图表，直观展示了不同环境因素下居民幸福值的分布情况。这些图表不仅能够帮助我们更清晰地认识民众福祉的实际情况，还能够为政策制定者提供有力的数据支持，以制定更加符合民众需求的政策措施。

数据背景：2021 年中国综合社会抽样调查

关键词：线性回归，方差分析，异方差修正

目录

摘 要.....	I
目录.....	II
第一章 绪论.....	1
1.1 研究背景及意义.....	1
1.2 研究现状.....	1
1.3 数据介绍.....	3
第二章 描述性统计分析.....	5
2.1 因变量：数值说明.....	5
2.2 抑郁情绪对幸福值的影响.....	6
2.3 不同学历对幸福值的影响.....	7
2.4 房屋面积对幸福值的影响.....	8
2.5 健康程度对幸福值的影响.....	9
2.6 对环境污染的感知对幸福值的影响.....	10
2.7 生活受极端天气影响对幸福值的影响.....	11
2.8 是否拥有汽车对幸福值的影响.....	12
2.9 是否有宗教信仰对幸福值的影响.....	13
2.10 不同收入水平对居民幸福感的影响.....	14
2.11 政治面貌对幸福值的影响.....	16
2.12 社交频率对幸福值的影响.....	18
第三章 回归分析.....	19
3.1 普通线性回归的结果与解读.....	19
3.2 变量相关性矩阵.....	20
第四章 自变量的方差分析.....	22
4.1 无交互作用的双因素方差分析：.....	22
4.1.1. 数据结构.....	22
4.1.2. 分析步骤.....	23
4.2 有交互作用的双因素方差分析：.....	25
4.3 方差分析.....	25
4.3.1 财产因素的方差分析.....	25
4.3.2 工作因素的方差分析.....	26

4.3.3 模型诊断—异方差.....	27
4.3.4 模型诊断—强影响点、正态性.....	28
参考文献.....	31
附录 A.....	32
附录 B.....	34

第一章 绪论

1.1 研究背景及意义

本文研究基于 2021 年中国综合社会抽样调查，旨在衡量个人生活质量、评估社会和谐与进步以及推动经济社会发展。随着我国经济的高速发展，人民生活水平不断提升，人们对幸福感、获得感和安全感的需求越来越强烈。习近平总书记在党的二十大报告中提出了“增进民生福祉，提高人民生活品质”的目标，着力为人民的幸福生活而努力。国民幸福值是衡量人们对自身生存体验和发展状况的一种指数，是评估社会进步的重要指标。因此，研究国民幸福值对制定更加精准有效的幸福发展战略具有重要意义。

国民幸福值是一种衡量人们对自身生存和发展状况感受与体验的指数，最早由南亚不丹王国的国王提出，他认为政策应关注幸福，并以实现幸福为目标。在人们的日常生活中，就业、治安、住房、医疗、教育等方面的改善对普通民众而言，更能带来幸福感，而国内生产总值或人均收入的增长只是其中之一。因此，国民幸福值相较 GDP 更具综合性和整体性，更全面和深刻地反映社会发展现状。

幸福值是对人们通常所说的幸福感的量化，它反映了人们根据自身价值标准对生活状态的满意度评价。这种评价涵盖了人们对生活总体以及主要生活领域的满意感、所体验到的快乐感，以及由于潜能实现而获得的价值感。作为检测社会良性运转的指标，幸福值更多地发挥一种诊断功能。通过追踪研究幸福值的变化，可以把握不同社会群体幸福值的走势和变化规律，为社会政策的制定和调整提供科学依据。幸福感指数的研究结果可以为政府决策提供参考。政府可以根据民众的幸福感受和需求，制定更加符合民意的政策，从而推动社会的和谐与稳定。

1.2 研究现状

目前，我国在国民幸福值方面取得了一定的成果。中国学术调查数据资料库发布的《幸福感数据集》分析了以十一个变量展开研究的调查数据，探讨了中国国民幸福值在各个维度的得分情况。中国社会科学院、中国人民大学等高校和研究机构也开展了相关研究，探讨了国民幸福值的测量方法和评估模型及影响中国国民幸福感的因素等。总

体来说，我国在国民幸福值的研究和关注上逐渐加强，未来还有很大的发展空间和必要性。

我国学者对国民幸福值的探索起步相对西方国家较晚，开始于 20 世纪 90 年代。刘伟教授(北京大学)和程国栋教授(中科院)提出的国民幸福值指标体系中包括政治自由、经济机会、社会机会、安全保障、文化价值观、环境保护等六类一级指标，并对其各一级指标中涉及的具体评估内容及标准进行了探讨；邢占军教授(社科院)领导的研究小组将幸福界定为人们对生活质量的一种主观感受，以此为基础提出了由知足充裕体验指数、心理健康体验指数、成长发展体验指数、社会信心体验指数、目标价值体验指数、自我接受体验指数、人际适应体验指数、身体健康体验指数、心态平衡体验指数、家庭氛围体验指数等 10 个一级指标构成的国民幸福值指标体系，并且设计了一套由 40 个项目组成的《中国民众主观幸福感量表》来分别开会研讨测量这些评估指标，但对具体的评估标准及评估方法没有做深入探讨；刘正山所设计的国民幸福值指标体系包括由开心程度、个人外向指数、人际关系指数等组成的主观指标体系和由基尼系数、支出结构、婚姻美满度、就业、生命的价格、游戏规则的正性等指标组成的客观指标体系，并主张通过访谈法、问卷调查法等统计调查方法来获取数据。

中国人民大学中国调查与数据中心主持的《中国综合社会调查（CGSS）》项目连续性学术调查项目，自 2003 年起，每年一次，对中国大陆各省市自治区 10000 多户家庭进行连续性横截面调查。2024 年的 CGSS 调查工作已全面展开，包括面对面访问、电话调查以及网络调查等多种调查模式，并已经回收了超过 2000 份有效问卷，数据仍在不断增长中。CGSS 在数据采集上实施了高灵活度的混合调查模式，该方法融合了面对面访问的深入性、电话调查的即时性以及网络调查的高效性与广泛性，显著提升了数据采集的效率与质量。

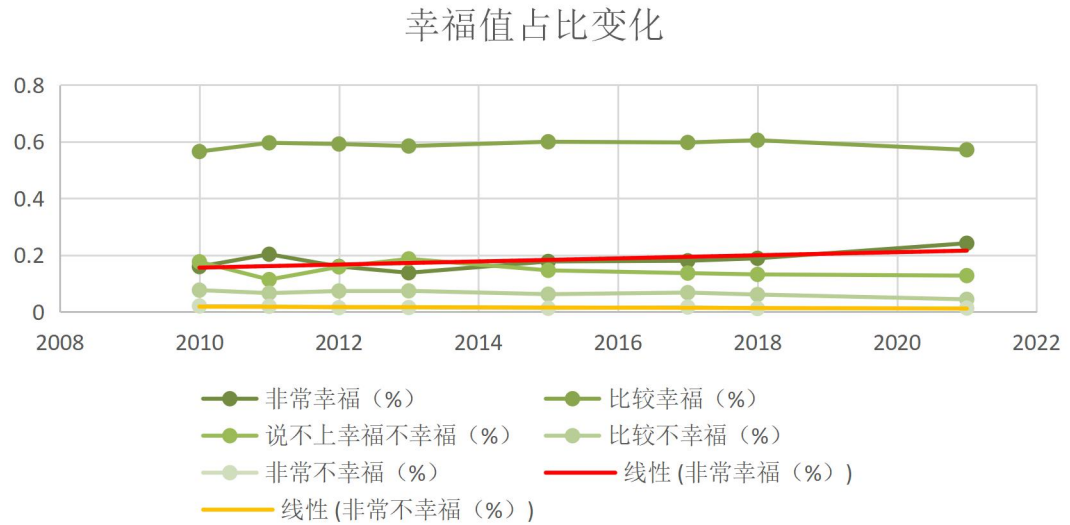
1.3 数据介绍

表 1 数据介绍

变量类型	变量名称	详细说明	取值范围
因变量	幸福值	共五个水平	非常不幸福、比较不幸福、说不上幸福不幸福、比较幸福、非常幸福
	抑郁情绪	共五个水平	总是、有时、经常、很少、从不
	社交	共七个水平	从来不、一年一次或更少、一年几次、大约一个月一次、一个月几次、一周 1-2 次、几乎每天
	工作满意度	共五个水平	非常不满意、不太满意、一般、比较满意、非常满意
	健康状况	共五个水平	很不健康、比较不健康、一般、比较健康、很健康
	公共服务态度	共四个水平	不满意、一般、满意、非常满意
	学历	共五个水平	没有受过任何教育、高中及以下、专科学历、本科学历、研究生及以上
	毕业情况	共两个水平	毕业、未毕业
	房屋面积	单位：平方米	10-500
	极端天气	共两个水平	受极端天气影响、不受极端天气影响
主观因素	生活环境污染	共五个水平	非常严重、比较严重、既严重也不严重、不太严重、根本不严重
	婚姻状况	共三个水平	已婚、离婚或丧偶及分居未离婚、未婚或同居
	政治面貌	共三个水平	共产党员、共青团员、群众
	个人收入	单位：元/年	0—9999999
	家庭收入	单位：元/年	0—9999999
	房产数量	共八个水平	0-8
	是否有车	共两个水平	是、否
	是否有宗教信仰	共两个水平	是、否

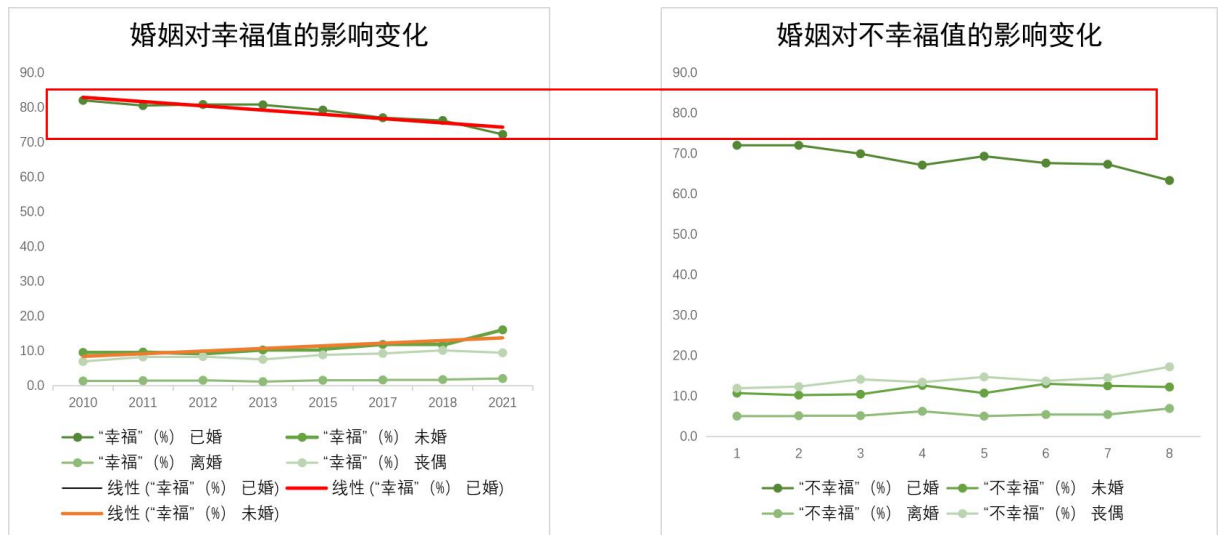
对于数据的处理，我们借鉴了天池数据集的相关研究，有针对的对问卷的不同维度进行提取，甄选出影响幸福值的主观因素和客观因素两大类，并逐一对影响幸福值的因素进行分析。

图 1 历年幸福值占比变化



从 2010 年到 2021 年的数据可以发现，我国“非常幸福”的人群以接近 50%的增速增至占比接近 24%；我国“非常不幸福”的人群以接近 38%的减速减少至 1.3%。

图 2 婚姻对幸福和不幸福值的影响



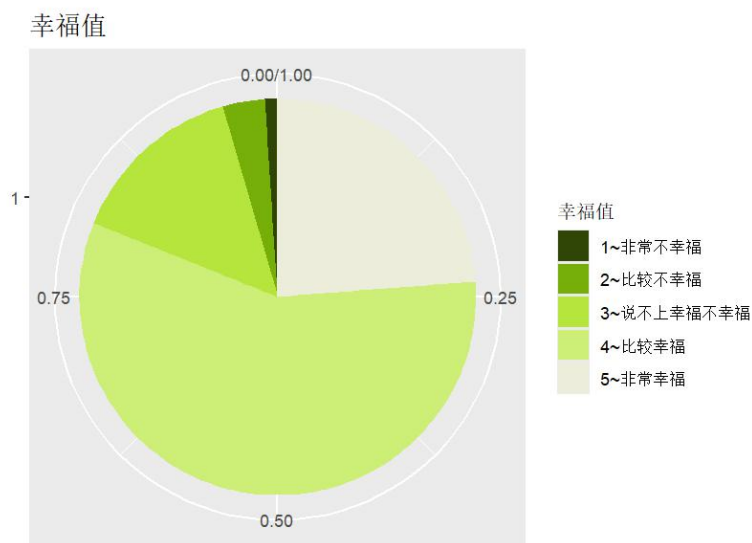
自 2010 年，认为婚姻能带来幸福感的人群以接近 1.19%的减速逐年减少，与此同时，未婚人群的幸福感逐年上升，并在 2021 年到达峰值 16%；相较而言，认为婚姻带来不幸福的人群也以接近 1.20 %的减速逐年减少，但总体占比少于认同婚姻能带来幸福感的人群。

第二章 描述性统计分析

2.1 因变量：数值说明

本文选择问卷中的“您现在是否感到幸福？”这一问题，将居民的答案划分为五个维度，依次分别为：“非常不幸福”，“比较不幸福”，“说不上幸福不幸福”，“比较幸福”，“非常幸福”，这五个维度，将成为本文探究的重点因变量。

图 3 因变量幸福值数值说明



在本研究的幸福感评估中，我们发现了一个显著的特征：处于“比较幸福”状态的人群占据了样本的最大比例，超过 50%。这一结果揭示了在当前社会环境下，大部分居民能够体验到一种相对较高的生活满意度，体现了整体社会福祉的积极态势。

进一步分析，我们发现“非常幸福”的人群比例也相当可观，接近总样本的四分之一。这一群体不仅在物质层面享有较好的生活条件，更在精神层面展现了高度的满足感和幸福感。他们的存在，为社会树立了追求更高生活质量与幸福感的典范。

然而，与此同时，我们也注意到幸福感分布的另一端。尽管“不幸福”状态的人群占比较少，但其中“非常不幸福”的群体仍然值得我们特别关注。这部分人群在生活满意度方面存在显著不足，可能面临着多重挑战与困境。他们的幸福感缺失，不仅影响了个人生活质量，也可能对社会稳定与和谐产生潜在影响。

表 2 幸福值的概括性度量指标

平均值	标准差	偏度系数	峰度系数
3.994	0.783	-0.935	1.698

幸福值的调查样本平均值为 3.994，这一数值意味着在所考察的群体中，所有个体的幸福感受的平均水平非常接近于 4 这一基准点。这为我们提供了一个关于整体幸福状况的大致概览，显示出大多数人在这方面的体验是趋于中等偏上水平的。

幸福值的标准差被计算为 0.783，这一数据为我们揭示了幸福值分布的一个重要特征：大多数个体的幸福感受与平均值 3.994 之间的偏差范围在 0.783 左右。这表明，虽然人们的幸福感受存在个体差异，但这些差异并没有过于悬殊，而是相对集中在一个较为适中的范围内。这样的标准差值说明幸福值的分布是比较均匀且稳定的。

进一步地，偏度系数为-0.935，这一负值向我们揭示了幸福值分布的另一个关键特点：分布呈现出向左偏斜的趋势。换句话说，虽然平均值接近 4，但大部分个体的幸福值实际上是集中在较高的水平线上，而有一些相对较少的个体则拥有较低的幸福值，这些较低的值在一定程度上拉低了整体的平均水平。这种偏斜分布表明，虽然大多数人感到相对幸福，但仍有一部分人群的幸福感受有待提升。

此外，峰度系数为 1.698，这一正值则揭示了幸福值分布的另一个现象：与正态分布相比，该分布具有更重的尾部或更尖的峰。这意味着在幸福值分布的两端，即极高和极低水平上的个体数量相对较多，而在平均水平附近的个体数量则相对较少。这种峰度特征进一步丰富了我们对于幸福值分布形态的理解，表明幸福感受在不同个体间存在较为显著的差异和多样性。

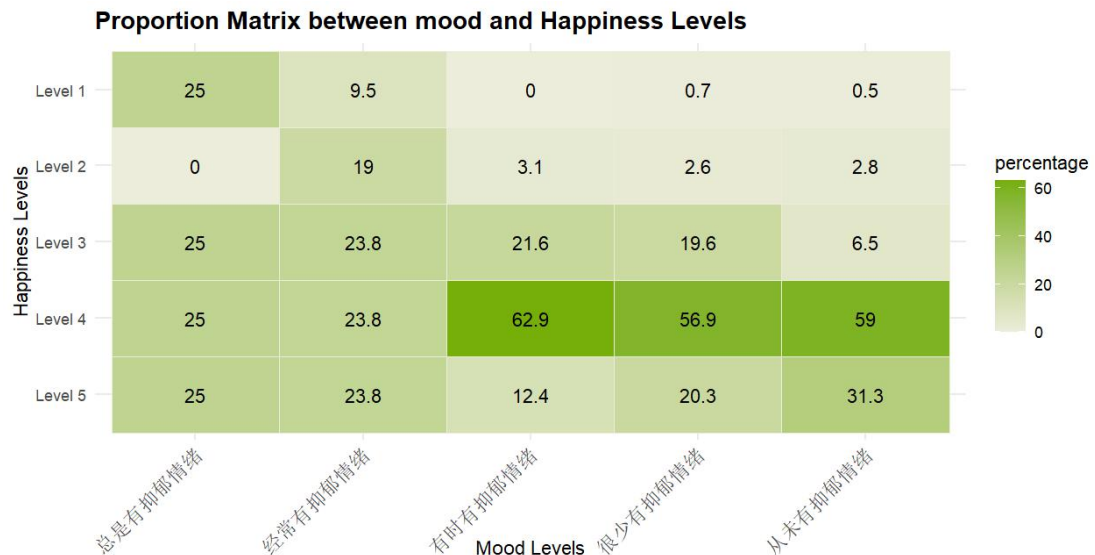
2.2 抑郁情绪对幸福值的影响

在深入探讨影响居民幸福值的诸多因素中，本研究特别选取了调查问卷中的一个关键问题：“您有抑郁情绪的频率是多少次？”作为核心自变量之一，以期更全面地理解心理健康状态对民众福祉的深远影响。为了精准衡量抑郁情绪的不同程度，我们将其频率细致地划分为五个阶段：“总是”、“经常”、“有时”、“很少”以及“从不”，并将这一变量视为五阶段的类别变量进行处理。

这样的分类不仅有助于我们细致地捕捉到抑郁情绪在不同人群中的分布情况，还能

深入探究其对居民幸福值产生的具体影响。通过系统地分析这一变量，我们期望能够揭示出抑郁情绪与居民福祉之间的潜在联系，从而为提升民众心理健康水平、促进整体福祉提供有力的科学依据和针对性的政策建议。

图 4 抑郁情绪对幸福值的影响



“很少、从未有过抑郁情绪”的人群所获得的幸福感最高，而“总是、经常有抑郁情绪”的人群所获得幸福值集中分布在 3。

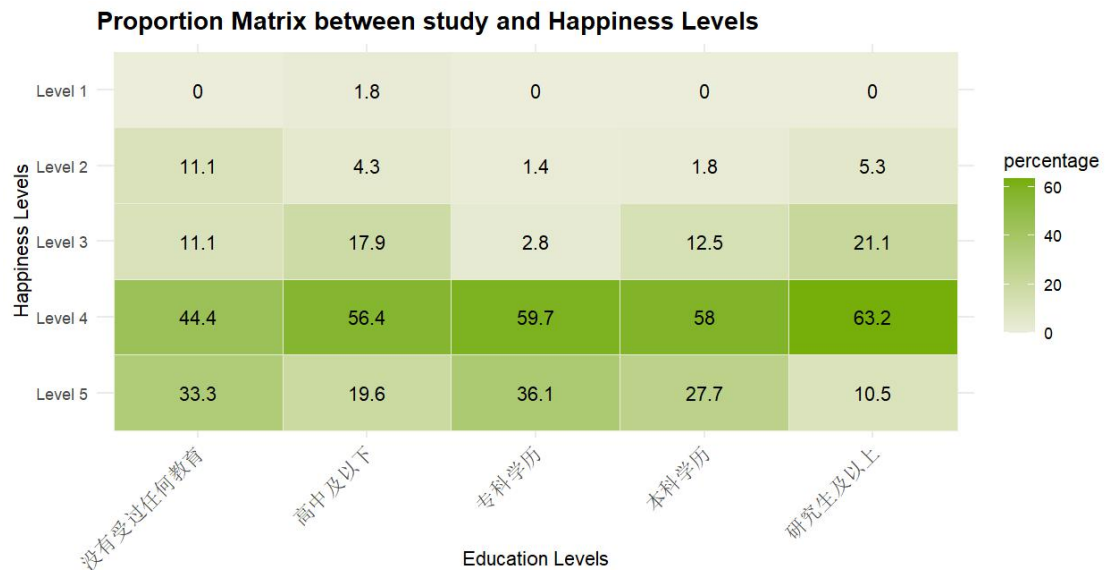
这一对比的结果，不仅加深了我们对抑郁情绪与幸福感之间复杂互动机制的理解，也提醒我们关注心理健康的重要性。它告诉我们，维护良好的心理状态，减少抑郁情绪的侵扰，是提升个人幸福感、促进生活品质的关键所在。因此，无论是从个人成长的角度出发，还是从社会和谐稳定的宏观层面考虑，都应加大对心理健康教育的投入，提升公众对心理健康问题的认识与重视程度，共同营造一个更加积极、健康的生活环境。

2.3 不同学历对幸福值的影响

将学历作为自变量研究幸福值，有助于深入探讨教育水平对个人幸福感的影响。这种研究可以揭示不同学历层次的个体在幸福感上的差异。

此外，这一研究对于评估教育政策的效果、指导教育资源的合理分配、促进社会公平和提高国民整体幸福感具有重要意义。通过了解学历与幸福值之间的关系，可以为制定更加有效的教育和社会政策提供科学依据，帮助政府和教育机构更好地满足人们对于教育和幸福感的需求。

图 5 学历对幸福值的影响



本研究发现，在幸福值为 4 和 5 的人群中（即较高的幸福值水平），占比最高的是专科学历和研究生学历。其余变量水平十分接近，这说明学历对幸福值的影响并不明显，这一发现为我们重新审视学历与幸福感之间的关系提供了新的视角。

它提示我们，尽管学历作为衡量个人知识水平和专业技能的一个重要指标，在职业发展、社会地位等方面可能发挥着重要作用，但在决定个体幸福感这一更为复杂且多维的领域里，其作用或许并不像人们普遍认为的那样绝对和单一。相反，幸福感的塑造可能更多地依赖于个人的心态、生活态度、人际关系以及社会支持网络等多重因素的共同作用。

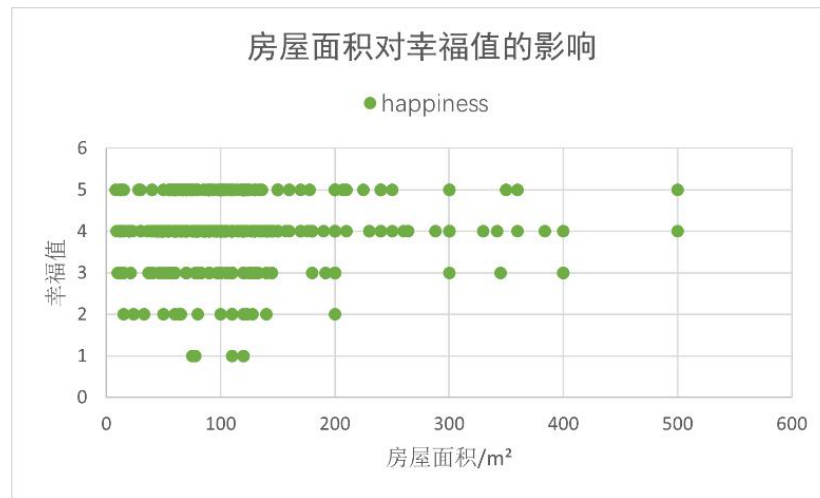
2.4 房屋面积对幸福值的影响

将房屋面积作为关键自变量深入研究其对居民幸福值的影响，这一举措不仅揭示了住房条件这一关键因素在塑造居民幸福感中的复杂作用机制，而且为政府及相关部门在制定和评估住房政策时提供了坚实而科学的依据。通过细致入微地考察不同房屋面积条件下居民幸福感的具体变化，政府能够得以窥见住房问题背后隐藏的深层次社会需求和民众心声，从而更加精准、有针对性地调整和优化住房政策，确保这些政策能够切实满足不同社会群体的多元化需求。

此外，这一研究结果同样为城市规划者带来了宝贵的信息资源和启示。它促使城市规划者在设计城市布局、构建居住环境时，不仅要考虑城市的经济发展和功能布局，更要将居民的幸福感纳入核心考量范畴，努力打造更加人性化、宜居、环保的城市环境，

让城市成为居民心中真正的幸福家园。

图 6 房屋面积对幸福值的影响

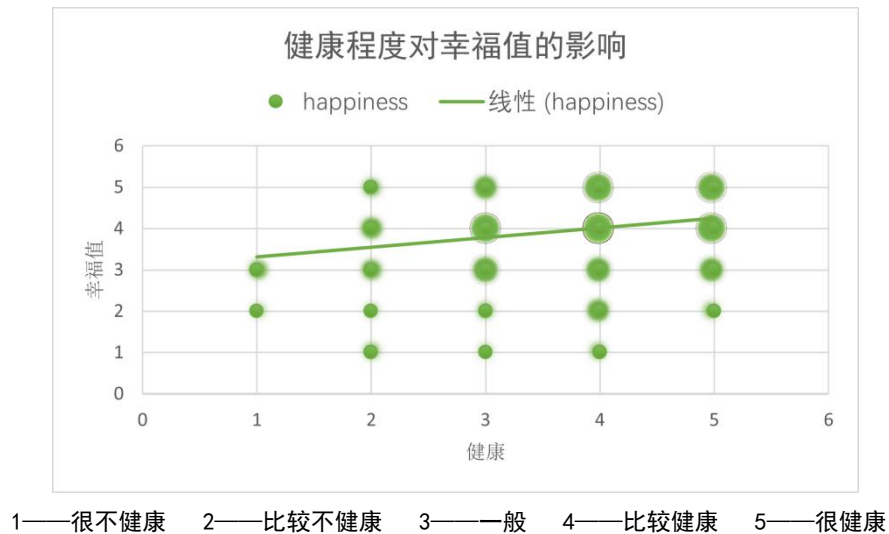


本研究发现，被调查的居民拥有的房屋面积集中分布在 100m^2 的左右，且此区间内幸福值集中在 3-5，处于较高的幸福值水平阶段；但当房屋面积超过 200m^2 时，可以明显看出随着房屋面积的增加，居民的幸福值会逐步提高，二者呈现正相关关系

2.5 健康程度对幸福值的影响

健康状况是影响个体幸福感的重要因素之一，我们通过对问卷中“您是否对自身的身体状况感到健康？”进行提取，将健康状态分为了五个阶段，依次探究其对幸福值的影响程度。通常情况下，良好的健康状况通常意味着较少的疾病和疼痛，更高的生活质量和活动能力，从而有助于提升个体的幸福感。

图 7 健康程度对幸福值的影响



在本研究的深入剖析中，我们发现健康程度与幸福感之间存在一种显著的正相关关系。具体而言，当健康程度被评定为最低等级 1 时，受访者的最高幸福感得分仅为 3。这一结果清晰地反映出，较差的健康状况对个体的幸福感有着明显的制约作用，限制了人们享受更高层次生活质量的潜力。

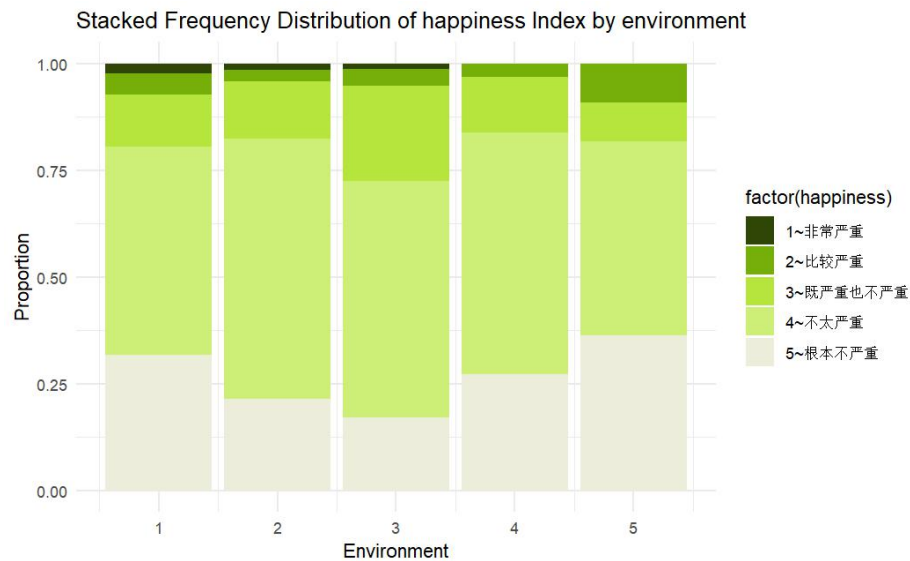
然而，随着健康程度的提升，幸福感的分布特征发生了显著变化。当健康程度达到 4 和 5 时，受访者的幸福值主要集中于 4 和 5 的高分区间。这一发现不仅揭示了健康状况对幸福感提升的积极影响，还进一步强调了保持身心健康对于提升生活质量、增强幸福感的重要性。

通过对数据的细致分析，我们绘制出了健康程度与幸福感之间的线性关系图。该图直观地展示了幸福值随着健康程度的增大而呈现出递增的趋势。这一线性关系不仅强化了健康与幸福感之间的正相关联系，还为我们提供了一个有力的依据，即改善健康状况是提升幸福感的有效途径之一。

2.6 对环境污染的感知对幸福值的影响

将环境污染程度的感知作为自变量研究幸福值具有多方面的意义。它不仅揭示了环境与幸福感之间的复杂关系，还为政策制定、城市规划、环境保护意识提升以及心理学研究提供了重要的参考和依据。

图 8 对环境污染的感知对幸福值的影响



当个体对环境污染的感知较低，甚至认为其并未产生任何影响时，其幸福值达到 5 的比例显著增大。这一结果反映了在相对清洁、健康的生活环境中，人们更容易获得较高的幸福感。

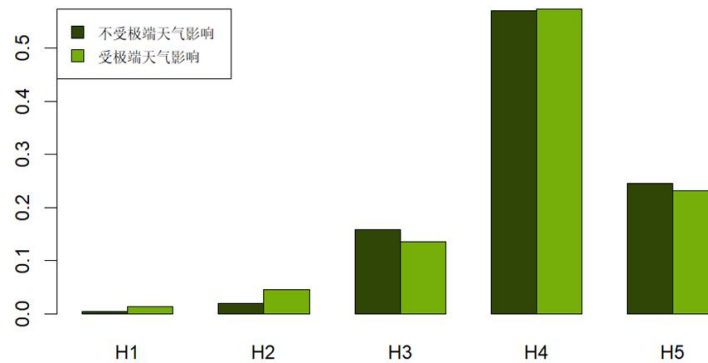
进一步分析发现，当个体对环境污染的感知逐渐增强，认为其问题较为严重时，低幸福感的情况开始频繁出现。这一趋势表明，环境污染的加剧不仅威胁着人们的身体健康，还可能对个体的心理状态产生负面影响，进而降低其幸福感。

值得注意的是，环境污染与幸福感之间的关系并非单向的。环境污染的严重程度是影响幸福感的一个重要因素，但同时，个体对环境污染的感知和态度也在一定程度上塑造着他们的幸福感体验。因此，在探讨环境污染对幸福感的影响时，我们需要综合考虑环境污染的客观状况以及个体对其的主观感知。

2.7 生活受极端天气影响对幸福值的影响

通过将极端天气作为自变量，可以系统地研究其对人类幸福感的具体影响，从而为应对气候变化提供科学依据。极端天气可能导致焦虑、抑郁等心理健康问题，进而影响个人的幸福感。通过研究极端天气与幸福感的关系，可以为心理健康服务提供新的视角，帮助制定更有效的干预措施。

图 9 生活受极端天气影响对幸福值的影响



当个体的幸福值处于较低水平时，报告其生活受到极端天气显著影响的占比相对较高。这一发现初步表明，在幸福感较低的群体中，极端天气可能作为一个额外的压力源，加剧了其生活的不确定性和挑战性，从而进一步降低了幸福感。

然而，当我们深入分析数据时，一个更为细致且复杂的图景逐渐浮现。通过对比不同幸福值水平下，生活受极端天气影响程度的柱状图，我们发现了一个令人惊讶的结果：尽管在低幸福感群体中，受极端天气影响的比例较高，但从整体比例来看，极端天气对幸福值的影响似乎并不具备显著的差异性。换句话说，无论是在高幸福感还是低幸福感的群体中，极端天气对生活的影响比例都相对接近，没有表现出明显的幸福值依赖性。

这一发现挑战了我们对极端天气与幸福感之间关系的传统认知，提示我们极端天气的影响可能更多地是作为一种普遍性的外部因素存在，而非特定针对幸福感高低的人群。它强调了极端天气作为一个全球性的挑战，对所有社会成员都可能构成潜在威胁，而不论其当前的幸福感水平如何。

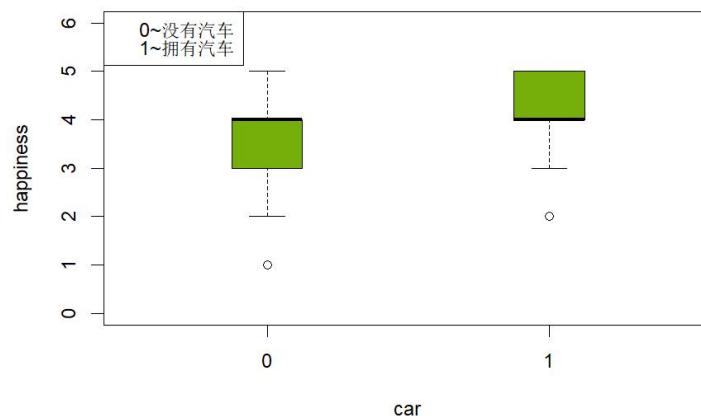
它提醒我们，在应对极端天气挑战时，需要采取更加全面和包容的策略，确保所有社会成员都能得到充分的保护和支持，从而维护并提升整体的幸福感水平。

2.8 是否拥有汽车对幸福值的影响

是否拥有汽车：拥有汽车可能表示较好的经济条件，但也可能带来经济压力，如购车贷款、保养费用等。汽车提供了出行的便利，特别是在公共交通不发达的地区，有助于提高生活满意度。汽车在某些社会和文化中也是地位的象征，可能影响个人的社交圈和幸福感。

汽车在某些社会和文化中也是地位的象征，可能影响个人的社交圈和幸福感。虽然汽车提供了出行的自由，但也可能带来安全和维护方面的担忧。汽车可能促进社交活动，如自驾游、探访亲友等，但也可能因交通拥堵等问题影响心情。虽然汽车提供了出行便利，但长时间驾驶可能对健康产生负面影响，如久坐不动导致的健康问题。

图 10 拥有汽车对幸福值的影响



在探究汽车拥有量与个体幸福感之间潜在联系的过程中，本研究借助箱线图这一工具，揭示了两者间的一个有趣且显著的关联。具体而言，通过对不同汽车拥有情况下的幸福感数据进行可视化呈现，我们发现拥有汽车的群体普遍在幸福感分布上展现出更高的水平。

箱线图清晰地展示了拥有汽车与未拥有汽车两组人群在幸福感得分上的差异。对于拥有汽车的群体而言，其幸福感的中位数明显偏高，且数据分布相对集中，表明这一群体在整体上享有较高的生活满意度。相比之下，未拥有汽车的群体在幸福感得分上则呈现出更为分散的分布，且中位数相对较低，反映了这部分人群在幸福感体验上的多样性和相对较低的平均水平。

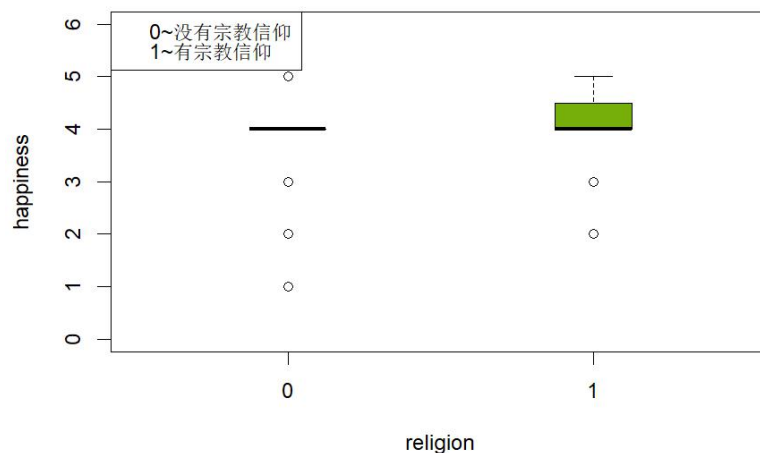
这一发现不仅揭示了汽车作为现代生活的一部分，对于提升个体幸福感可能具有的积极作用，也提示我们汽车拥有量可能作为衡量社会经济地位和生活质量的一个重要指标，与幸福感之间存在着密切的关联。

2.9 是否有宗教信仰对幸福值的影响

宗教信仰：宗教信仰作为研究幸福值的一个方面具有重要意义。它不仅能够提供心

灵寄托和希望、引导道德伦理和行为规范、形成社交网络和共同体，还能够促进心理健康和内部和平、提升生活满意度和幸福感，并有助于进行跨文化比较和理解。然而，需要注意的是，宗教信仰与幸福感之间的关系可能受到多种因素的影响，包括个人的宗教地位、文化背景、社会经济状况等。因此，在研究宗教信仰与幸福感的关系时，需要综合考虑这些因素的作用。

图 11 宗教信仰对幸福值的影响



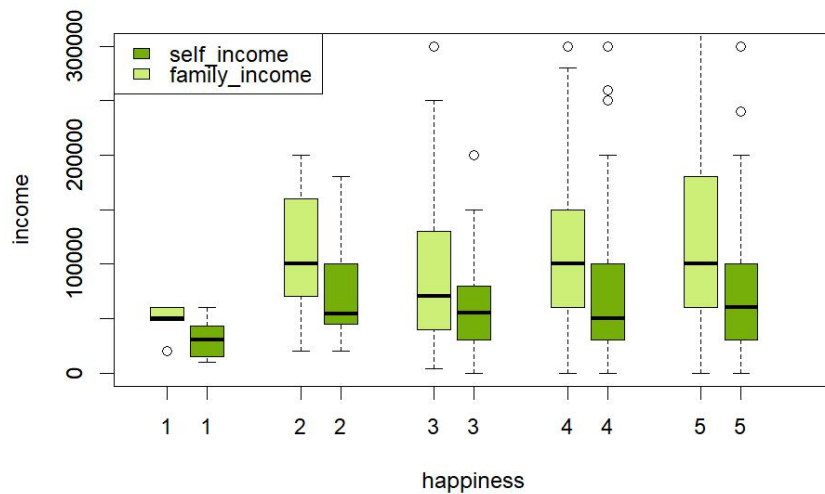
相较于“有宗教信仰”的群体，“无宗教信仰”的群体在幸福值上呈现出更为集中的分布特征。这一结果表明，在无宗教信仰的人群中，幸福感水平相对统一，个体差异较小，可能反映了这一群体在价值观、生活态度等方面存在较为一致的认知。

然而，在“有宗教信仰”的群体中，幸福感的分布则呈现出更为广泛的态势。这一发现不仅揭示了宗教信仰对个体幸福感影响的复杂性，还进一步提示我们，宗教信仰对每个人的影响程度和方式可能存在显著差异。具体而言，在“有宗教信仰”的人群中，幸福感的高低差异较大，这可能与个体对宗教信仰的理解、信仰程度、信仰实践以及信仰所带来的心理支持和社会归属感等因素密切相关。

2.10 不同收入水平对居民幸福感的影响

收入水平：收入是经济状况的直接反映，物质基础决定幸福感，家庭收入是衡量一个家庭经济状况的最直观指标。较高的收入通常意味着更多的物质财富和更高的生活水平，从而有助于提升个体的幸福感。它不仅直接反映了人们的物质生活条件和经济压力状况，还与社会比较、未来预期以及家庭关系和社会稳定等因素密切相关。

图 12 不同收入水平对居民幸福感的影响 1



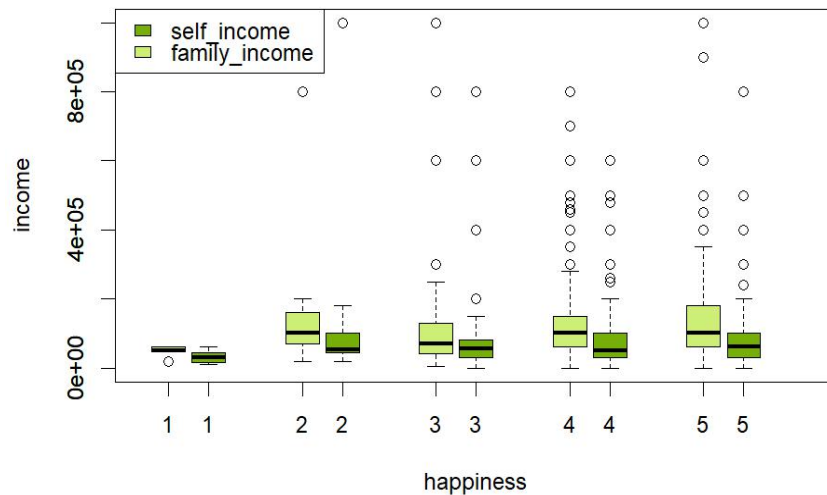
在研究收入水平与个体幸福感之间关系的过程中，本研究发现：幸福值更高的人群，其家庭收入水平的中位数普遍高于幸福值较低的人群。这一结果不仅进一步证实了收入作为影响幸福感的关键因素之一，其具体作用机制也在不同幸福值水平的人群中得到了清晰的体现。

具体而言，通过对比分析，我们发现无论是个人收入还是家庭收入，都对幸福值为1（即最低幸福感水平）的人群产生了较为显著的影响。在个人收入层面，较低的收入水平往往限制了个体的生活选择和满足感，从而增加了其处于低幸福感状态的风险。而在家庭收入层面，较低的家庭收入可能导致家庭成员面临更大的经济压力和不确定性，进一步削弱了整体的幸福感。

值得注意的是，这种收入与幸福感之间的关系并非线性，而是呈现出一种复杂的非线性特征。也就是说，虽然收入水平的提升能够在一定程度上提高个体的幸福感，但当收入达到一定水平后，其对幸福感的边际贡献将逐渐减弱。此外，个体对收入的感知和期望也在很大程度上影响着幸福感水平，这解释了为何在相同收入水平下，不同个体的幸福感可能存在显著差异。

这一发现不仅为我们理解幸福感形成的复杂性提供了有益的线索，也提醒我们在制定相关政策、提升公众幸福感时，应充分考虑收入分配的公平性和个体的经济需求。

图 13 不同收入水平对居民幸福感的影响 2



为了更细致地探究收入水平与幸福感之间的关系，我们对原始箱线图进行了缩放处理，将其尺寸缩小至原来的十分之一。

在缩小后的箱线图中，我们注意到一个有趣的现象：幸福值较高的人群，其收入水平向高收入的离散程度相较于其他幸福值水平的人群更高。具体而言，这部分人群的收入水平不仅中位数较高，而且数据的分布范围也更广，显示出更大的变异性。这一发现表明，在幸福值较高的人群中，收入水平的差异可能更加显著，且高收入个体在其中的占比相对较高。

这一结果进一步丰富了我们对于收入水平与幸福感之间关系的理解，即：虽然收入水平是影响幸福感的重要因素之一，但在不同幸福值水平的人群中，其影响方式和程度可能存在差异。在幸福值较高的人群中，高收入可能作为提升幸福感的一个重要因素，但其作用并非绝对，因为即使在相同的收入水平下，个体的幸福感也可能因其他因素（如家庭关系、健康状况、个人心态等）的不同而有所差异。

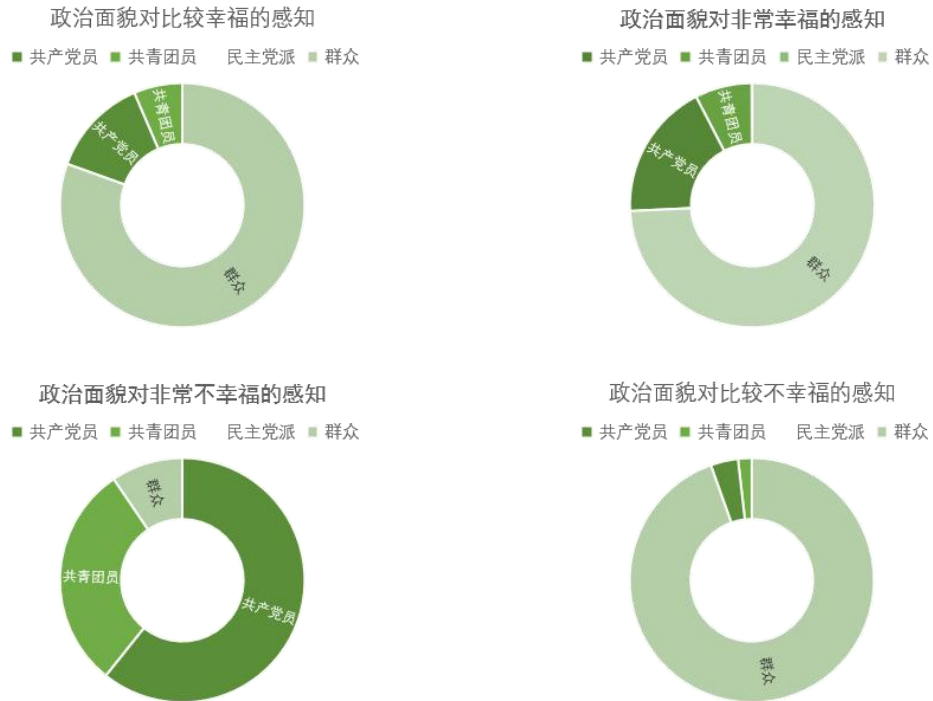
这一发现不仅为我们提供了关于幸福感形成机制的更多线索，也提醒我们在研究幸福感时，应充分考虑多种因素的共同作用，以及不同人群之间的差异性。

2.11 政治面貌对幸福值的影响

政治面貌：政治面貌可能反映了个体在社会政治生活中的参与程度和归属感，进而对幸福感产生影响。拥有中共党员、民主党派或共青团员等政治身份的人群可能更感到

幸福，因为这些身份可能为他们提供了更多的社会认同感和自我价值实现的成就感。

图 14 政治面貌对幸福值的影响



在探究幸福感感知的群体差异性时，我们发现了一个值得深思的现象：群众对于幸福感的感知是相当明显的，他们能够清晰地表达自己的幸福感水平。然而，与之形成鲜明对比的是，处于“非常不幸福”阶段的共产党员占比较大，这一结果无疑引发了我们对于共产党员幸福感状态的深切关注。

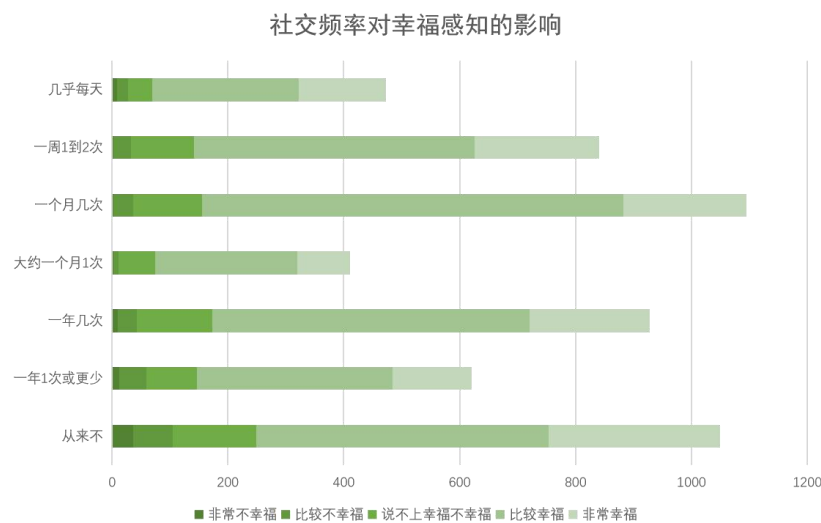
通过对比分析，我们初步推断，共产党员之所以在“非常不幸福”阶段占比较大，可能与他们肩负的责任和使命的特殊性密切相关。作为社会的先锋队和人民群众的引路人，共产党员不仅需要承担日常工作中的职责，还需要时刻铭记为人民服务的宗旨，这种高度的责任感和使命感无疑给他们带来了更大的心理压力和期望挑战。在这样的背景下，即使他们在物质层面取得了相对不错的成就，也可能因为精神层面的重负而难以感受到真正的幸福。

这一发现不仅为我们提供了关于幸福感形成机制的更多线索，也提醒我们在关注社会整体幸福感的同时，应更加关注特定群体的幸福感状态，为他们的身心健康和幸福感提升提供更多的支持和帮助。

2.12 社交频率对幸福值的影响

社交频率：社交活动是人们日常生活中不可或缺的一部分，它为个体提供了情感支持、归属感和认同感。通过观察或询问个体的社交频率，研究者可以间接推断其幸福感水平，因为社交活跃度往往与幸福感正相关。社交频率不仅反映了个体的社交能力，还可能影响其心理健康状态。

图 15 社交频率对幸福值的影响



针对柱状图的分析，本研究发现幸福感较高的人群，其社交频率呈现出一种相对集中且适度的分布态势。具体而言，这部分人群的社交活动主要集中在“一个月几次”、“一周1到2次”以及“从来不社交”这三个区间内。这一结果似乎表明，对于幸福感较高的人群而言，保持一种既不过于频繁也过于稀疏的社交节奏，可能是他们获得幸福感的一个重要因素。

相比之下，幸福感较低的人群在社交频率上的分布则呈现出一种相对稀疏且低频的特征。他们的社交活动主要集中在“大约1个月1次”以及“一年1次或更少”这两个区间内。这一发现进一步证实了社交频率与幸福感之间的正相关关系，即社交活动的减少可能导致幸福感的降低。

值得注意的是，虽然“从来不社交”也出现在幸福感较高人群的社交频率分布中，但这并不意味着社交活动的完全缺失是提升幸福感的必要条件。相反，这可能反映了这部分人群在社交需求上的独特性和选择性，他们可能更加注重社交活动的质量和深度，而非数量。因此，在理解社交频率与幸福感的关系时，我们需要综合考虑社交活动的频率、质量以及个体对社交活动的需求和期望。

这一发现不仅为我们理解幸福感形成的复杂性提供了有益的线索，也提醒我们在日

常生活中，应注重培养健康、积极的社交习惯，以更好地提升个人的幸福感水平。

第三章 回归分析

3.1 普通线性回归的结果与解读

表 3 平台线性回归结果

变量	取值	回归系数	P 值	显著性水平
是否是党员		0.14305	0.027797	*
是否健康		0.11460	0.006405	**
是否抑郁		0.11628	0.000148	***
是否对自己的工作满意	以否为基准组 是	0.31407	1.54e-13	***
是否有车		0.19338	0.005381	**
是否结婚		-0.07193	0.107350	.
有几套房屋产权		0.10472	0.024974	*

在深入剖析幸福感这一复杂心理现象的过程中，本研究借助逐步回归分析的强大工具，揭示了多个关键变量对幸福感产生的显著影响。这些变量不仅涵盖了心理健康、身体健康、工作满意度等主观因素，还涉及物质条件、社会身份等客观方面，共同构建了一个多维度的幸福感影响因素框架。

首先，从心理健康的角度来看，“是否抑郁”这一变量对幸福感的影响尤为显著。回归分析结果显示，抑郁程度的不同水平（从“总是”到“从不”）与幸福感呈现出负相关关系，即抑郁程度每减轻一个等级，幸福值便会相应提升 0.116。这一发现强调了心理健康在幸福感形成中的核心作用，提示我们维护良好的心理状态对于提升幸福感至

关重要。

同时，工作满意度作为另一个重要的主观因素，其对幸福感的影响同样不容忽视。回归分析显示，对工作满意的人群相较于不满意者，其幸福值显著提高了 0.314。这一结果不仅反映了工作满意度在个体幸福感中的重要地位，也揭示了职业发展与个人幸福感之间的紧密联系。

在身体健康方面，“是否健康”这一变量同样对幸福感产生了显著影响。与健康人群相比，不健康者的幸福值降低了 0.114。这一发现进一步强调了身体健康作为幸福感基础的重要性，提醒我们要重视身体健康的维护，以预防疾病、提升生活质量。

此外，物质条件也是影响幸福感不可忽视的因素之一。回归分析结果显示，“是否有车”这一变量对幸福感具有显著的正向影响，有车组的幸福值比无车组高出 0.193。这一发现不仅体现了物质条件对幸福感的一定贡献，也反映了现代社会中交通工具对个人生活便捷性和幸福感的影响。

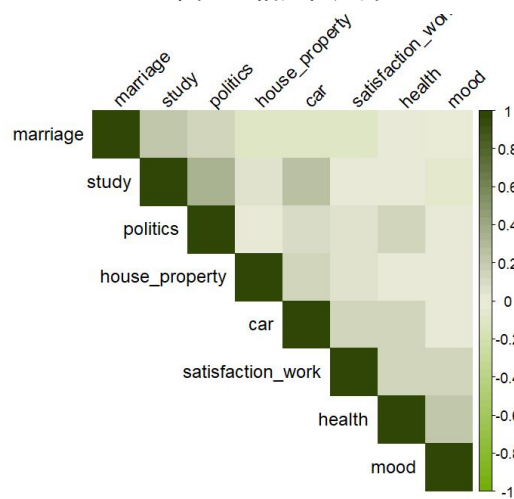
然而，值得注意的是，并非所有物质条件的改善都能带来幸福感的提升。例如，“已婚人群比未婚人群的幸福值低 0.071”这一结果，可能反映了婚姻状态对幸福感的复杂影响。这提示我们，在探讨物质条件与幸福感关系时，需要综合考虑个体需求、生活状态等多种因素。

最后，房屋产权作为衡量物质财富和社会地位的重要指标之一，也对幸福感产生了显著影响。回归分析显示，人们每增加拥有 1 套房，其幸福值便会提升 0.105。这一发现不仅揭示了房屋产权在提升个体安全感和社会地位方面的作用，也反映了住房条件对幸福感的重要贡献。

综上所述，本研究通过逐步回归分析，深入探索了多个关键变量对幸福感的影响，为我们理解幸福感的形成机制提供了丰富的实证依据。这些发现不仅有助于我们更全面地认识幸福感这一复杂心理现象，也为制定相关政策、提升公众幸福感提供了有益的参考。

3.2 变量相关性矩阵

图 16 相关性矩阵



通过逐步回归法，我们排除了变量之间的多重共线性，使得各个变量之间都体现出较好的独立性，经过变量间的相关性矩阵图验证，再一次说明这些变量的选取是相互独立且有效的，为模型的拟合结果，提供了较好的支撑。

第四章 自变量的方差分析

双因素方差分析（Two-Way ANOVA）是一种强大的统计方法，用于研究两个独立分类变量（也称为因素）对一个连续因变量的影响，以及这两个因素之间是否存在交互作用。

4.1 无交互作用的双因素方差分析：

4.1.1 数据结构

表 4 双因素方差分析的数据结构

	A	B	C	D	E	F	G
1			列因素(j)				平均值
2			列 1	列 2	...	列 r	$\bar{x}_{i.}$
3	行因素(i)	行 1	x_{11}	x_{12}	...	x_{1r}	$\bar{x}_{1.}$
4		行 2	x_{21}	x_{22}	...	x_{2r}	$\bar{x}_{2.}$
5		⋮	⋮	⋮	⋮	⋮	⋮
6		行 k	x_{k1}	x_{k2}	...	x_{kr}	$\bar{x}_{k.}$
7	平均值 $\bar{x}_{.j}$		$\bar{x}_{.1}$	$\bar{x}_{.2}$...	$\bar{x}_{.r}$	$\bar{\bar{x}}$

这 $k \times r$ 个总体中的每一个总体都服从正态分布，且有相同的方差。
表中， $\bar{x}_{i.}$ 是行因素的第 i 个水平下各观测值的平均值，其计算公式为

$$\bar{x}_{i.} = \frac{\sum_{j=1}^r x_{ij}}{r} \quad i = 1, 2, \dots, k$$

表中， $\bar{x}_{.j}$ 是列因素的第 j 个水平下各观测值的平均值，其计算公式为：

$$\bar{x}_{.j} = \frac{\sum_{i=1}^k x_{ij}}{k} \quad j = 1, 2, \dots, r$$

$\bar{\bar{x}}$ 是全部 kr 个样本数据的总平均值，其计算公式为：

$$\bar{\bar{x}} = \frac{\sum_{i=1}^k \sum_{j=1}^r x_{ij}}{kr}$$

4.1.2. 分析步骤

(1) 提出假设。

为了检验两个因素的影响，需要对两个因素分别提出如下假设。

对行因素提出的假设为：

$H_0: \mu_1 = \mu_2 = \dots = \mu_i = \dots = \mu_k$ 行因素（自变量）对因变量没有显著影响

$H_1: \mu_i (i = 1, 2, \dots, k)$ 不全相等 行因素（自变量）对因变量有显著影响

对列因素提出的假设为：

$H_0: \mu_1 = \mu_2 = \dots = \mu_j = \dots = \mu_r$ 列因素（自变量）对因变量没有显著影响

$H_1: \mu_j (j = 1, 2, \dots, r)$ 不全相等 列因素（自变量）对因变量有显著影响

均值。

(2) 构造检验统计量。

为检验 H_0 是否成立，需要分别确定检验行因素和列因素的统计量。与单因素方差分析构造统计量的方法一样，这里也需要从总平方和的分解入手。总平方和是全部样本观察值 x_{ij} ($i=1, 2, \dots, k; j=1, 2, \dots, r$) 与总的样本平均值元的误差平方和，记为 SST，即

$$\begin{aligned} SST &= \sum_{i=1}^k \sum_{j=1}^r (x_{ij} - \bar{x})^2 \\ &= \sum_{i=1}^k \sum_{j=1}^r (x_{i.} - \bar{x})^2 + \sum_{i=1}^k \sum_{j=1}^r (x_{.j} - \bar{x})^2 + \sum_{i=1}^k \sum_{j=1}^r (x_{ij} - x_{i.} - x_{.j} + \bar{x})^2 \end{aligned}$$

其中，分解后的等式右边的第一项是行因素所产生的误差平方和，记为 SSR，即

$$SSR = \sum_{i=1}^k \sum_{j=1}^r (x_{i.} - \bar{x})^2$$

第二项是列因素所产生的误差平方和，记为 SSC，即第二项是列因素所产生的误差平方和，记为 SSC，即

$$SSC = \sum_{i=1}^k \sum_{j=1}^r (x_{.j} - \bar{x})^2$$

第三项是除行因素和列因素之外的剩余因素所产生的误差平方和，称为随机误差平方和，记为 SSE，即

$$SSE = \sum_{i=1}^k \sum_{j=1}^r (x_{ij} - x_{i.} - x_{.j} + \bar{x})^2$$

上述各平方和的关系为：

$$SST = SSR + SSC + SSE$$

在上述误差平方和的基础上计算均方，也就是将各平方和除以相应的自由度。与各误差平方和相对应的自由度分别是：

总平方和 SST 的自由度为 $kr - 1$ ；

行因素的误差平方和 SSR 的自由度为 $k - 1$ ；

列因素的误差平方和 SSC 的自由度为 $r - 1$ ；

随机误差平方和 SSE 的自由度为 $(k - 1)(r - 1)$ 。

为构造检验统计量，需要计算下列各均方：

行因素的均方，记为 MSR，即

$$MSR = \frac{SSR}{k - 1}$$

列因素的均方，记为 MSC，即

$$MSC = \frac{SSC}{r - 1}$$

随机误差项的均方，记为 MSE，即

$$MSE = \frac{SSE}{(k - 1)(r - 1)}$$

为检验行因素对因变量的影响是否显著，采用下面的统计量：

$$F_R = \frac{MSR}{MSE} \sim F(k - 1, (k - 1)(r - 1))$$

为检验列因素的影响是否显著，采用下面的统计量：

$$F_C = \frac{MSC}{MSE} \sim F(r - 1, (k - 1)(r - 1))$$

(3) 作出统计决策。

计算出检验统计量后，根据给定的显著性水平 α 和两个自由度，查 F 分布表得到相应的临界值 F_α ，然后将 F_R 和 F_C 与 F_α 进行比较。

若 $F_R > F_\alpha$ ，则拒绝原假设 $H_0: \mu_1 = \mu_2 = \dots = \mu_i = \dots = \mu_k$ ，表明 $\mu_i (i = 1, 2, \dots, k)$ 之间的差异是显著的；也就是说，所检验的行因素对观测值有显著影响。

若 $F_C > F_\alpha$ ，则拒绝原假设 $H_0: \mu_1 = \mu_2 = \dots = \mu_j = \dots = \mu_r$ ，表明 $\mu_j (j = 1, 2, \dots, r)$ ，

之间的差异是显著的；也就是说，所检验的列因素对观测值有显著影响。

为使计算过程更加清晰，通常将上述过程的内容列成方差分析表，其一般形式如表 2 所示。

表 5 双因素方差分析表

	A	B	C	D	E	F	G
1	误差来源	误差平方和 SS	自由度 df	均方 MS	F 值	P 值	F 临界值
2	行因素	SSR	k-1	MSR	F_R		
3	列因素	SSC	r-1	MSC	F_C		
4	误差	SSE	(k-1) × (r-1)	MSE			
5	总和	SST	kr-1				

4.2 有交互作用的双因素方差分析：

如果两个因素搭配在一起会对因变量产生一种新的效应，就需要考虑交互作用对因变量的影响，这就是有交互作用的双因素方差分析。提出假设时，需要对行变量、列变量和交互作用变量分别提出假设。

基于对数据的研究分析，作者将从两个方面对人们的幸福值展开方差分析；分别为：财产因素和社会关系因素。

4.3 方差分析

4.3.1 财产因素的方差分析

表 6 财产因素对于幸福值的影响

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
house	1	0.7	0.740	0.740	0.248	
car	1	29.9	29.890	54.015	2.67e-13	***
house:car	1	1.1	1.065	1.925	0.165	
Residuals	2535	1402.8	0.553			

本文选择了房、车作为财产因素项目的两大构成类目，分析发现：房和车之间的交互作用不显著，意味着我们可以分别考虑它们对幸福值的影响，而不需要考虑它们之间的相互作用；需要注意的是，车这个因素的 P 值非常小（ $2.67e-13$ ），远远小于 0.05 的显著性水平，因此我们可以确信车因素对幸福值有显著影响。

在进一步探讨财产因素与幸福值的关系时，我们可以发现车作为现代生活中不可或缺的交通工具，其拥有与否往往成为衡量个人或家庭财产状况的重要指标。此次分析结果显示，车因素对幸福值具有显著的正向影响，这可能是因为车辆的拥有能够带来出行的便利与舒适，进而提升生活品质与幸福感。

相比之下，房因素虽然也是人们普遍关注的财产之一，但在此分析中其对幸福值的影响并不显著。这或许是因为房产的拥有更多关联于居住需求与安全感的满足，而非直接提升幸福感。此外，房产价值的波动较大，受经济环境、政策调整等多重因素影响，因此其对幸福值的贡献可能相对复杂且多变。

4.3.2 工作因素的方差分析

对工作满意程度的高低，以及居民的政治面貌特征，这两者之间极有可能存在着一种微妙的交互作用机制。具体而言，居民的政治面貌不仅是一个身份标识，更可能在一定程度上深刻地影响着他们的职业选择、工作的时间安排、工作的地点偏好等一系列客观因素。这些由政治面貌所带来的职业路径差异，进一步地，可能会直接或间接地作用于居民对工作满意程度的感知与评价。

例如，拥有特定政治面貌的居民可能更容易获得某些特定行业或职位的青睐，这些岗位或许能提供更好的职业发展前景、更稳定的工作保障或是更高的社会地位，从而增强他们对工作的满意度。反之，若政治面貌限制了某些职业机会，可能导致居民在职业选择上受限，进而影响到他们对工作的整体满意度。因此，在探讨工作满意程度时，不可忽视居民政治面貌这一潜在的重要变量，以及它可能带来的复杂交互效应。

表 7 工作因素对于幸福值的影响

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Satisfaction_work	1	42.68	42.68	82.620	2e-16	***
Politics	1	3.56	3.56	6.889	0.00894	**
Satisfaction_work:Politics	1	2.68	2.68	5.188	0.02317	*
Residuals	488	252.07	0.52			

本文选择了对工作是否满意、政治面貌作为工作因素项目的两大构成类目，分析发现：对工作的满意程度和政治面貌之间的交互作用显著，意味着我们不可以分别考虑它们对幸福值的影响；需要注意的是，政治面貌因素的 P 值非常小（0.00894），远远小于 0.05 的显著性水平，因此我们可以确信政治面貌因素对幸福值有显著影响。

4.3.3 模型诊断--异方差

表 8 B-P 检验结果

Studentized Breusch-Pagan test	
BP	18.679
df	8
p-value	0.01667

利用 BP 检验，对该线性回归模型进行异方差检验，发现该模型的 P 值为 0.01，小于显著性水平，故认为该模型存在异方差问题。利用异方差稳健标准误的方法对该模型进行修正，得到的结果如下：

表 9 异方差稳健标准误修正结果

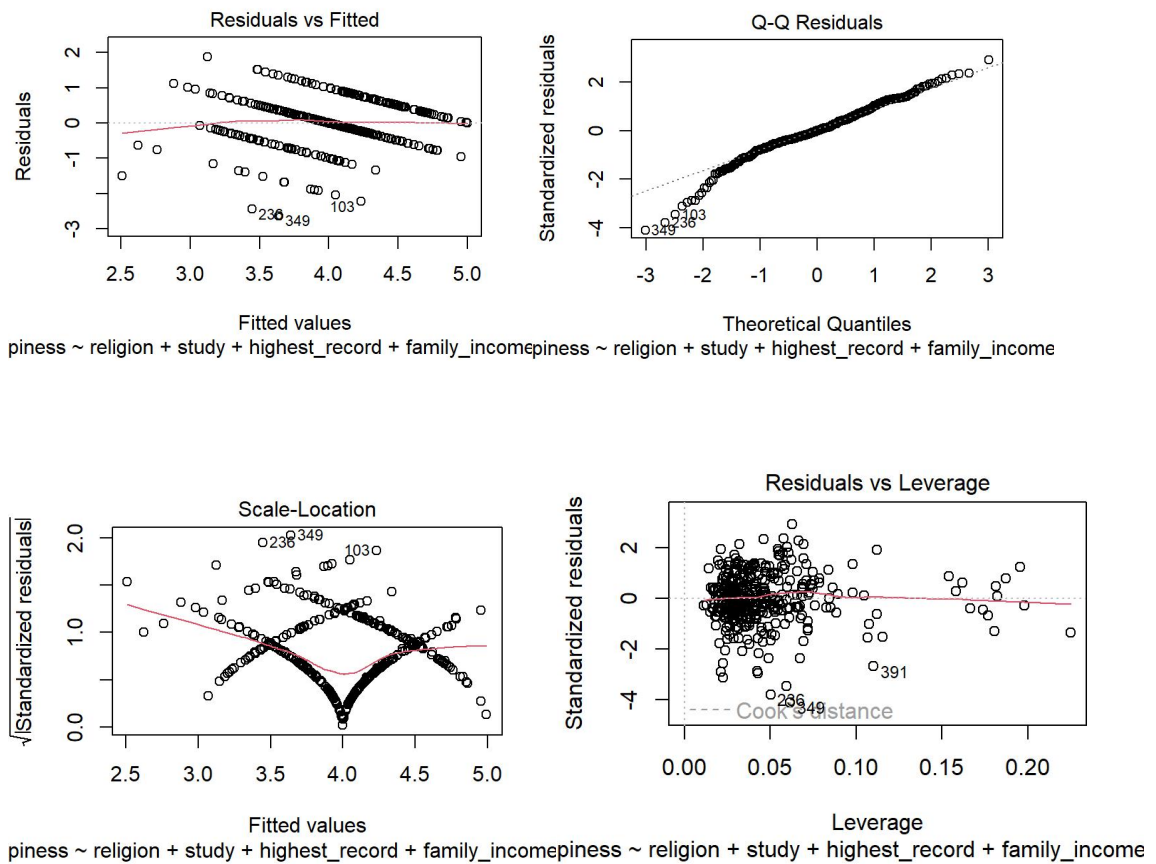
T test of coefficients					
	Estimate	Std. Error	t value	Pr(> t)	显著 水平
politics	0.156502	0.060085	2.6047	0.0095518	**
health	0.112896	0.051077	2.2103	0.0276689	*
mood	0.118029	0.033588	3.5140	0.0004938	***
satisfaction_work	0.315085	0.047918	6.5755	1.578e-10	***
car	0.201258	0.071507	2.8145	0.0051354	**
marriage	0.216533	0.086002	2.5178	0.0122144	*
environment	-0.052562	0.035143	-1.4956	0.1355624	
Multiple R-squared	0.2761				
Adjusted R-squared	0.263				
Residual standard error:	0.6596 on 386 degrees of freedom				

对该模型进行异方差稳健标准误修正后，得到的结果显著性水平较高；修正后的 R 方拟合效果一般，仍需继续改进。

4.3.4 模型诊断—强影响点、正态性

对于建立好的线性回归模型，我们针对其模型拟合的效果进行进一步的检验，对模型的效果进行进一步诊断。

图 17 线性模型诊断分析



针对左上图，即“Residuals vs Fitted”图（残差与拟合值图），我们观察到残差与估计值之间存在着一种较为明显的负相关关系。这种关系的存在，意味着当前的模型在预测不同水平的数据时，其预测误差（即残差）并非随机分布，而是呈现出一种系统性的模式。这通常表明模型在捕捉数据中的真实关系方面还有一定的不足，尚有改进的空间。这种负相关关系也从一个侧面解释了R方（决定系数）所反映的拟合程度较低的现象，即模型对于数据的整体拟合效果不够理想，尚有一部分数据变异未能被模型有效解释。

针对右上图，即“Normal QQ图”，这张图揭示了因变量数据在正态性方面的表现。从图中我们可以看出，数据点大致沿着参考线（即对角线）分布，这说明了该因变量在一定程度上符合正态分布的特性。正态分布是线性回归分析中一个重要的前提假设，因为它保证了回归分析的统计推断（如参数估计和假设检验）的有效性和可靠性。因此，该因变量的正态性为我们进行线性回归分析建模提供了良好的前提基础。

再来看左下图，即“Scale-Location图”（尺度-位置图）。在这张图中，我们注意到拟合曲线呈现出一种微笑曲线的状态，这是异方差问题的一个典型表现，与上述的

B-P 检验效果一致，异方差意味着不同观测值的残差具有不同的方差，违反了线性回归分析中关于误差项同方差的假设。异方差问题的存在会导致参数估计的标准误被低估，从而影响统计推断的准确性。因此，我们需要进一步对模型进行修正，以解决这一异方差问题。

最后，针对右下图，即“Residuals vs Leverage 图”（残差与杠杆值图），我们观察到所有的观测点都较为集中地分布在图的中心区域，没有出现特殊的强影响点。强影响点是指那些对模型参数估计和预测结果产生较大影响的观测值，它们可能是由于数据录入错误、异常值或测量误差等原因产生的。在这张图中，没有观测点显示出异常的杠杆值或残差，这表明我们的数据中没有明显的强影响点存在，这对于模型的稳定性和可靠性是一个积极的信号。

综上所述，通过对这四张诊断图的综合分析，我们可以对有针对性的认识到模型目前存在的问题，从而为后续的模型改进和优化提供有力的依据。

参考文献

- [1]邢占军. 幸福指数的指标体系构建与追踪研究[J]. 2006(8):10—12.
- [2]周四军, 庄成杰. 基于距离综合评价法的我国国民幸福指数 NHI 测评[J]. 财经理论与实践, 2008:9.
- [3]王紫博, 刘凯歌, 赵雲. 中国国民幸福指数统计测度与影响因素分析[J]. 中国商论, 2023(20):034-038.
- [4]罗新阳. 解读幸福指数 [J] . 宁波广播电视大学学报, 2007, 5(1): 97 — 101.
- [5]何艳桃, 马文省. 国民幸福指数研究现状及反思[J]. 渭南师范学院报, 2012. 09.
- [6]朝克, 何浩, 李佳钰. 国民幸福指数评价体系构建及实证[J]. 统计观察, 2016. 04.

附录 A

表 10 居民幸福感分布情况及变化趋势

	2010	2011	2012	2013	2015	2017	2018	2021
非常幸福 (%)	0.159853829	0.203420021	0.160269533	0.138137083	0.178033416	0.18039965	0.189007203	0.242107195
比较幸福 (%)	0.565309765	0.596009975	0.59177755	0.584622144	0.599744362	0.597245442	0.605151895	0.571402349
说不上幸福 (%)	0.176850514	0.114178839	0.158990106	0.187170475	0.146626495	0.136852161	0.132320702	0.128120411
比较不幸福 (%)	0.076910003	0.06644104	0.073865575	0.074428822	0.062631243	0.068306663	0.061384278	0.044603524
非常不幸福 (%)	0.02107589	0.019950125	0.015097236	0.015641476	0.012964485	0.017196083	0.012135922	0.01376652
样本数	11767	5614	11724	11380	10953	12561	12772	5448
“幸福”占比 (%)	0.725163593	0.799429996	0.752047083	0.722759227	0.777777778	0.777645092	0.79415909	0.813509545
“不幸福”占比 (%)	0.097985893	0.086391165	0.088962811	0.090070299	0.075595727	0.085502747	0.0735202	0.058370044

表 11 居民幸福感分布情况及变化趋势（以健康状况和婚姻状况划分）

		2010	2011	2012	2013	2015	2017	2018	2021
“幸福” (%)	比较不健康	12.0	33.7	12.9	11.4	12.4	13.6	13.8	11.4
	比较健康	35.7	24.7	38.2	39.3	41.0	37.7	41.0	38
	很不健康	2.4	8.9	2.7	1.9	2.1	3.4	2.5	3.9
	很健康	28.5	9.1	22.7	29.3	24.1	20.5	21.0	20.4
	一般	21.1	22.0	23.2	17.7	20.1	24.5	21.5	26
“不幸” (%)	比较不健康	28.2	36.9	31.2	25.2	32.0	30.2	32.2	30.1
	比较健康	22.9	12.7	23.1	26.2	24.2	20.2	23.9	18.5
	很不健康	11.7	30.5	11.4	9.6	13.0	14.9	13.8	20.4
	很健康	12.4	4.5	11.2	16.0	9.9	8.5	8.3	9.1
	一般	24.8	14.4	23.0	22.8	20.6	25.8	21.5	21.6
“幸福” (%)	已婚	82.0	80.5	80.8	80.7	79.2	77.0	76.2	72.2
	未婚	9.5	9.6	9.1	10.2	10.4	11.8	11.6	16.0
	离婚	1.3	1.4	1.5	1.1	1.5	1.6	1.7	2
	丧偶	6.9	8.2	8.3	7.5	8.8	9.2	10.1	9.4
	已婚	72.0	72.0	69.9	67.1	69.3	67.6	67.3	63.3
“不幸” (%)	未婚	10.7	10.2	10.4	12.6	10.7	13.0	12.5	12.2
	离婚	5.0	5.1	5.1	6.2	5.0	5.4	5.4	6.9
	丧偶	11.9	12.3	14.1	13.4	14.7	13.7	14.5	17.2

附录 B

```
#对数据预处理

library(readxl)

happy <- read_excel("C:/Users/HUAWEI/Desktop/数据 1(2).xlsx")


#print("Original DataFrame:")

#print(df)

#print(happy)

#删除序号

happy <- happy[, -1]

happy <- happy[, -8]

happy <- happy[, -20]#是否幸福总的来说

happy <- happy[, -21]#健康，总的来说

happy <- happy[, -23]

happy <- happy[, -23]

happy <- happy[, -7]

happy <- happy[, -19]

happy <- happy[, -19]

happy <- happy[, -18]

happy <- happy[, -11]

happy <- happy[, -11]


#重命名

colnames(happy)<-c("religion", "study", "highest_record", "personal_income", "politics",

                  "house_square", "health", "mood", "social", "happiness", "satisfaction_work",

                  "family_income", "house_property", "car", "marriage",

                  "environment", "extreme_weather")


##分类变量处理
```

#宗教信息处理

```
happy$religion <- ifelse(happy$religion == "从来没有参加过", 0, 1)
```

#学历信息处理

#删除该列中没有选择学历变量的数据

```
happy <- happy[!grepl("\\.", happy$study), ]
```

```
library(dplyr)
```

```
happy <- happy %>%
```

```
  mutate(study = case_when(
```

```
    study == "没有受过任何教育" ~ 1,
```

```
    study == "技校" ~ 2,
```

```
    study == "私塾、扫盲班" ~ 2,
```

```
    study == "小学" ~ 2,
```

```
    study == "中专" ~ 2,
```

```
    study == "初中" ~ 2,
```

```
    study == "普通高中" ~ 2,
```

```
    study == "职业高中" ~ 2,
```

```
    study == "大学专科（成人高等教育）" ~ 3,
```

```
    study == "大学专科（正规高等教育）" ~ 3,
```

```
    study == "大学本科（正规高等教育）" ~ 4,
```

```
    study == "大学本科（成人高等教育）" ~ 4,
```

```
    study == "研究生及以上" ~ 5,
```

```
    TRUE ~ NA_real_
```

```
  ))
```

#毕业情况处理（分为毕业和未毕业的情况）

```
happy$highest_record <- ifelse(happy$highest_record == "毕业", 1, 0)
```

#个人收入处理(国家统计局《2023 年国民经济和社会发展统计公报》)

```
exclude_values <- c("不知道", "不适用", "拒绝回答")
```

```
happy <- happy[!(happy$personal_income %in% exclude_values), ]

library(dplyr)

happy <- happy %>%

  mutate(personal_income = case_when(

    personal_income <= 9215 ~ 1, # 低收入

    personal_income <= 20442 ~ 2, # 注意：这里会自动排除<=9215的情况，因为前面的条件已经匹配

    personal_income <= 32195 ~ 3, # 同理

    personal_income <= 50220 ~ 4,

    personal_income <= 95055 ~ 5,

    personal_income > 95055 ~ 6, # 超高收入，注意这里使用>而不是<=

    TRUE ~ NA_real_

  ))
```

#家庭收入处理(国家统计局《2023年国民经济和社会发展统计公报》)

```
exclude_values <- c("不知道", "不适用", "拒绝回答")

happy <- happy[!(happy$family_income %in% exclude_values), ]

happy <- happy %>%

  mutate(family_income = case_when(

    family_income <= 9215 ~ 1, # 低收入

    family_income <= 20442 ~ 2, # 中低收入

    family_income <= 32195 ~ 3, # 中等收入

    family_income <= 50220 ~ 4, # 中高收入

    family_income <= 95055 ~ 5, # 高收入人群

    family_income > 95055 ~ 6, # 超高收入

    TRUE ~ NA_real_

  ))
```

#社会身份处理

```
exclude_values <- c("不知道", "不适用", "拒绝回答")

happy <- happy[!(happy$politics %in% exclude_values), ]
```



```
library(dplyr)
```

```
happy <- happy %>%
```

```
  mutate(politics = case_when(
    politics == "群众" ~ 1,
    politics == "共产党员" ~ 2,
    politics == "共青团员" ~ 3,
    TRUE ~ NA_real_
  ))
```

```
#房屋面积处理
```

```
library(dplyr) #《国务院办公厅转发建设部等部门关于做好稳定住房价格工作意见的通知》（国办发〔2005〕26号）
```

```
happy <- happy %>%
```

```
  filter(!(house_square == "不知道" | house_square == "拒绝回答"))
```

```
happy <- happy %>%
```

```
  mutate(house_square = case_when(
    house_square <= 120 ~ 1, #普通住宅
    house_square <= 144 ~ 2, #中高档住宅
    house_square > 144 ~ 3, #高档住宅
    TRUE ~ NA_real_
  ))
```

```
#房屋产权信息
```

```
library(dplyr)
```

```
happy <- happy %>%
```

```
  filter(!(house_property == "拒绝回答"))
```

```
#健康信息处理
```

```
library(dplyr)
```

```
happy <- happy %>%
```

```
mutate(health = case_when(

  health == "很不健康" ~ 1,

  health == "比较不健康" ~ 2,

  health == "一般" ~ 3,

  health == "比较健康" ~ 4,

  health == "很健康" ~ 5,

  TRUE ~ NA_real_

))

happy <- happy %>%

  filter(!(health == "不知道"))

#抑郁情绪处理

library(dplyr)

happy <- happy%>%

  mutate(mood = case_when(

    mood == "总是" ~ 1,

    mood == "有时" ~ 2,

    mood == "经常" ~ 3,

    mood == "很少" ~ 4,

    mood == "从不" ~ 5,

    TRUE ~ NA_real_

  ))

happy <- happy %>%

  filter(!(mood == "不知道" | mood == "拒绝回答"))

#幸福值的处理

happy <- happy[!grepl("\\.", happy$happiness), ]

happy <- happy %>%

  filter(!(happiness == "不知道" | happiness == "拒绝回答"))

library(dplyr)
```

```
happy <- happy%>%

mutate(happiness = case_when(

  happiness == "非常不幸福" ~ 1,

  happiness == "比较不幸福" ~ 2,

  happiness == "说不上幸福不幸福" ~ 3,

  happiness == "比较幸福" ~ 4,

  happiness == "非常幸福" ~ 5,

  TRUE ~ NA_real_

))

#社交数据处理

happy <- happy[!grepl("\\.", happy$social), ]

happy <- happy %>%

  filter(!(social == "不知道" ) | social == "拒绝回答")

library(dplyr)

happy <- happy%>%

  mutate(social = case_when(

    social == "从来不" ~ 1,

    social == "一年 1 次或更少" ~ 2,

    social == "一年几次" ~ 3,

    social == "大约一个月 1 次" ~ 4,

    social == "一个月几次" ~ 5,

    social == "一周 1 到 2 次" ~ 6,

    social == "几乎每天" ~ 7,

    TRUE ~ NA_real_

  ))

#车数据处理

happy <- happy %>%

  filter(!(car == "不知道" ))
```

```
library(dplyr)

happy <- happy%>%

  mutate(car = case_when(

    car == "有" ~ 1,

    car == "没有" ~ 0,

    TRUE ~ NA_real_

  ))

#环境数据处理

happy <- happy[!grepl("\\.", happy$environment), ]

happy <- happy %>%

  filter(!(environment == "无法选择"))

library(dplyr)

happy <- happy%>%

  mutate(environment = case_when(

    environment == "非常严重" ~ 1,

    environment == "比较严重" ~ 2,

    environment == "既严重也不严重" ~ 3,

    environment == "不太严重" ~ 4,

    environment == "根本不严重" ~ 5,

    TRUE ~ NA_real_

  ))

#婚姻状况处理

library(dplyr)#这里自定义为: 已婚, 未婚, 离婚, 丧偶

happy <- happy%>%

  mutate(marriage = case_when(

    marriage == "初婚有配偶" ~ 1,

    marriage == "再婚有配偶" ~ 1,

    marriage == "分居未离婚" ~ 0,
```

```
marriage == "离婚" ~ 0,

marriage == "丧偶" ~ 0,

marriage == "同居" ~ 0,

marriage == "未婚" ~ 0,

TRUE ~ NA_real_

))

#极端天气数据处理

happy <- happy %>%#这里自定义为：有影响，无影响

  filter(!(extreme_weather == "无法选择"))

library(dplyr)

happy <- happy %>%

  mutate(extreme_weather = case_when(

    extreme_weather == "根本没有影响" ~ 0,

    extreme_weather == "有一点影响" ~ 1,

    extreme_weather == "有一些影响" ~ 1,

    extreme_weather == "有很大影响" ~ 1,

    extreme_weather == "有非常大的影响" ~ 1,

    TRUE ~ NA_real_

  ))

#工作满意情况处理

happy <- happy[!grepl("\\.", happy$satisfaction_work), ]

happy <- happy %>%

  filter(!(satisfaction_work == "不知道"))

happy <- happy %>%

  mutate(satisfaction_work = case_when(

    satisfaction_work == "非常不满意" ~ 1,

    satisfaction_work == "不太满意" ~ 2,

    satisfaction_work == "一般" ~ 3,
```

```

satisfaction_work == "比较满意" ~ 4,

satisfaction_work == "非常满意" ~ 5,

TRUE ~ NA_real_

))

#####

#描述性统计分析（政治面貌）

library(ggplot2)

ggplot(happy, aes(x = social, y = happiness)) +

  geom_bar(stat = "identity") +

  scale_x_discrete(labels = c("从来不", "一年 1 次或更少", "一年几次", "大约一个月 1 次", "一个月几次", "一周 1
到 2 次", "几乎每天")) +

  labs(x = "Custom Category", y = "Value") +

  theme_minimal()

#党员身份

count_table <- table(happy$politics, happy$happiness)

politics_df <- data.frame(happy$politics, happy$happiness)

data$percent <- ave(politics_df$count, politics_df$happy.happiness, FUN = function(x) x / sum(x) * 100)

df_with_count <- politics_df%>%

  group_by(happy.politics, happy.happiness) %>%

  mutate(count = n()) %>%

  ungroup()

#####

#删除有缺失值的行

happy <- na.omit(happy)

# 绘制环境污染程度段不同幸福感指数的堆积频数分布直方图

count_happiness <- happy %>% group_by(environment, happiness) %>% summarise(Count = n())

```

```

count_happiness

custom_palette<-c("#304606","#76af0a","#b6e63e","#cdee77","#eceddb")

library(ggplot2)

h<-ggplot(happy, aes(x = environment, fill = factor(happiness))) +

  geom_bar(position = "fill") + # 使用填充位置来创建堆积条形图

  labs(x = "Environment", y = "Proportion", title = "Stacked Frequency Distribution of happiness Index by
environment") +

  theme_minimal() +

  scale_fill_manual(values=custom_palette,labels=c("1~非常严重","2~比较严重","3~既严重也不严重","4~不太严重
","5~根本不严重"))# 使用 Set1 调色板

h

# 绘制天气情况对不同幸福感指数的堆积频数分布直方图

library(gmodels)

weather_table <- CrossTable(happy$extreme_weather, happy$happiness)

weather_data <- as.matrix(data.frame("H1" = c(0.005, 0.014), "H2" = c( 0.020, 0.045 ),

                                     "H3" = c(0.158, 0.135), "H4" = c(0.571, 0.574), "H5" = c(0.246, 0.232)))

barplot(weather_data, col = c("#304606","#76af0a"), beside = TRUE)

#图例

legend("topleft",

       legend = c("不受极端天气影响", "受极端天气影响"),

       fill = c("#304606","#76af0a"),cex=0.8 )

## 相关性矩阵矩阵图

happy1 <- happy[,c("study", "politics", "health", "mood", "satisfaction_work", "house_property", "car", "marriage")]

library(corrplot)

happy1<-na.omit(happy1)

descriptive$offertype <- as.factor(ifelse(descriptive$offertype=="Admitted", T, F)) # 调整为因子变量

cor_happy <- cor(happy1)

corrplot(cor_happy, method = "color",type = "upper", order = "hclust", tl.col = "black", tl.srt = 45)

corrplot(cor_happy, method = "color",col = colorRampPalette(c("#76af0a","#eceddb","#304606"))(50),type = "upper",

```

```

order = "hclust", tl.col = "black", tl.srt = 45)

# 绘制收入分析箱线图

#箱线图

# 检查并转换列为数值型，如果无法转换则返回 NA

happy$personal_income<- as.numeric(as.character(happy$personal_income))

happy$family_income<- as.numeric(as.character(happy$family_income))

# 过滤掉包含 NA 值的行

happy<- happy[!is.na(happy$family_income), ]

# 打印处理后的数据框

print("Data Frame after removing non-numeric rows:")

print(df)

boxplot(personal_income~happiness, data = happy,boxwex = 0.25,at =1:5+0.15, col="#76af0a",
xlab="happiness",ylab="income", ylim=c(0,300000))

boxplot(family_income~happiness, data = happy, add = TRUE, boxwex = 0.25,at = 1:5 - 0.15, col="#cdee77")

legend("topleft", c("self_income", "family_income"),fill=c("#76af0a","#cdee77"))

boxplot(happiness~religion, data = happy,boxwex = 0.25, col="#76af0a", xlab="religion",ylab="happiness",
ylim=c(0,6))

legend("topleft",

      legend = c("0~没有宗教信仰", "1~有宗教信仰"),

      cex=1 )

boxplot(happiness~car, data = happy,boxwex = 0.25, col="#76af0a", xlab="car",ylab="happiness", ylim=c(0,6))

legend("topleft",

      legend = c("0~没有汽车", "1~拥有汽车"),

      cex=1 )

```



```
##因变量饼图

library(ggplot2)

piechar1<-ggplot(happy,aes(x=factor(1),fill=factor(happiness)))+

  geom_bar(position = "fill",width = 2)+scale_fill_manual("幸福值
",values=c("#304606","#76af0a","#b6e63e","#cdee77","#eceddb"),labels=c("1~非常不幸福","2~比较不幸福","3~说不上幸
福不幸福","4~比较幸福","5~非常幸福"))+

  coord_polar(theta = "y")+labs(x="",y="",title = "\n 幸福值")

piechar1

#抑郁情绪与幸福感的矩阵图

library(reshape2)

freq_table <- table(happy$mood, happy$happiness)

print(freq_table)

freq_table_long <- melt(freq_table)

colnames(freq_table_long) <- c("mood", "happiness", "frequency")

freq_table_long$mood <- factor(freq_table_long$mood, levels = 1:5, labels = c("总是有抑郁情绪", "经常有抑郁情绪",
"有时有抑郁情绪", "很少有抑郁情绪", "从未有抑郁情绪"))

freq_table_long$happiness <- factor(freq_table_long$happiness, levels = 1:5, labels = paste("Level", 1:5))

ggplot(freq_table_long, aes(x = mood, y = happiness, fill = frequency)) +

  geom_tile(color = "#eceddb") +

  scale_fill_gradient(low = "#eceddb", high = "#76af0a") +

  theme_minimal() +

  geom_text(aes(label = ifelse(frequency > 0, frequency, NA)), color = "black") +

  labs(title = "Frequency Matrix between Mood and Happiness Levels", x = "Mood", y = "Happiness Levels") +

  theme(axis.text.x = element_text(angle = 45, hjust = 1), plot.title = element_text(size = 14, face = "bold")) +

  scale_y_discrete(limits = rev(levels(freq_table_long$happiness)))

prop_table <- prop.table(freq_table, margin = 1) * 100

prop_table_long <- melt(prop_table)

colnames(prop_table_long) <- c("mood", "happiness", "percentage")
```

```

prop_table_long$mood <- factor(prop_table_long$mood, levels = 1:5, labels = c("总是有抑郁情绪", "经常有抑郁情绪", "有时有抑郁情绪", "很少有抑郁情绪", "从未有抑郁情绪"))

prop_table_long$happiness <- factor(prop_table_long$happiness, levels = 1:5, labels = paste("Level", 1:5))

ggplot(prop_table_long, aes(x = mood, y = happiness, fill = percentage)) +

  geom_tile(color = "#eceddb") +

  scale_fill_gradient(low = "#eceddb", high = "#76af0a") +

  theme_minimal() +

  geom_text(aes(label = round(percentage, 1)), color = "black") +

  labs(title = "Proportion Matrix between mood and Happiness Levels", x = "Mood Levels", y = "Happiness Levels") +

  theme(axis.text.x = element_text(angle = 45, hjust = 1), plot.title = element_text(size = 14, face = "bold")) +

  scale_y_discrete(limits = rev(levels(prop_table_long$happiness)))

#学历与幸福感的矩阵图

library(reshape2)

freq_table <- table(happy$study, happy$happiness)

print(freq_table)

freq_table_long <- melt(freq_table)

colnames(freq_table_long) <- c("study", "happiness", "frequency")

freq_table_long$study <- factor(freq_table_long$study, levels = 1:5, labels = c("没有受过任何教育", "高中及以下", "专科学历", "本科学历", "研究生及以上"))

freq_table_long$happiness <- factor(freq_table_long$happiness, levels = 1:5, labels = paste("Level", 1:5))

ggplot(freq_table_long, aes(x = study, y = happiness, fill = frequency)) +

  geom_tile(color = "#eceddb") +

  scale_fill_gradient(low = "#eceddb", high = "#76af0a") +

  theme_minimal() +

  geom_text(aes(label = ifelse(frequency > 0, frequency, NA)), color = "black") +

  labs(title = "Frequency Matrix between Study and Happiness Levels", x = "Study", y = "Happiness Levels") +

  theme(axis.text.x = element_text(angle = 45, hjust = 1), plot.title = element_text(size = 14, face = "bold")) +

  scale_y_discrete(limits = rev(levels(freq_table_long$happiness)))

```

```

prop_table <- prop.table(freq_table, margin = 1) * 100

prop_table_long <- melt(prop_table)

colnames(prop_table_long) <- c("study", "happiness", "percentage")

prop_table_long$study <- factor(prop_table_long$study, levels = 1:5, labels = c("没有受过任何教育", "高中及以下", "
专科学历", "本科学历", "研究生及以上"))

prop_table_long$happiness <- factor(prop_table_long$happiness, levels = 1:5, labels = paste("Level", 1:5))

ggplot(prop_table_long, aes(x = study, y = happiness, fill = percentage)) +

  geom_tile(color = "#eceddb") +

  scale_fill_gradient(low = "#eceddb", high = "#76af0a") +

  theme_minimal() +

  geom_text(aes(label = round(percentage, 1)), color = "black") +

  labs(title = "Proportion Matrix between study and Happiness Levels", x = "Education Levels", y = "Happiness
Levels") +

  theme(axis.text.x = element_text(angle = 45, hjust = 1), plot.title = element_text(size = 14, face = "bold")) +

  scale_y_discrete(limits = rev(levels(prop_table_long$happiness)))

#####

#线性回归模型

#确认训练集和预测集

library(ggplot2)

library(lattice)

library(caret)

set.seed(123) # 设置种子以保证结果可重复

# 创建训练集的索引（80%的数据用于训练）

happy <- na.omit(happy)

trainIndex <- createDataPartition(happy$happiness, p = 0.8, list = FALSE)

# 划分为训练集和测试集

```

```
trainData <- happy[trainIndex, ] # 训练集

testData <- happy[-trainIndex, ] # 测试集


mymodel <- lm(happiness ~ religion + study + highest_record + family_income + politics +
              house_square + health + mood + social + satisfaction_work +
              house_property + car + marriage + environment + extreme_weather, data = trainData)

summary(mymodel)


#检验多重共线性

library(carData)

library(car)

vif(mymodel)


# 计算数据框中所有变量的相关性矩阵

x <- data.frame(happy$religion , happy$study , happy$highest_record , happy$family_income , happy$politics ,
               happy$house_square , happy$health , happy$mood , happy$social , happy$satisfaction_work ,
               happy$house_property , happy$car , happy$marriage , happy$environment ,
happy$extreme_weather)

cor_matrix <- cor(x)

print(cor_matrix)


#去除多重共线性

mymodel <- step(mymodel, direction = "both")

summary(mymodel)


#####


#方差分析

anova_model_estate <- aov(happiness ~ satisfaction_work * politics, data = happy)

summary(anova_model_estate)
```

```
anova_model_estate1 <- aov(happiness ~ satisfaction_work * mood, data = happy)

summary(anova_model_estate1)

anova_model_estate <- aov(happiness ~ health * marriage, data = happy)

summary(anova_model_estate)

#用 ln 修正

mymodel_ln <- lm(log(happiness) ~ religion + study + highest_record + family_income + politics +
                  house_square + health + mood + social + satisfaction_work +
                  house_property + car + marriage + environment + extreme_weather, data = trainData)

summary(mymodel_ln)

library(lmtest)

bptest(mymodel)

#减轻了异方差，但依然存在异方差，尝试异方差稳健标准误修正

#针对 mymodel_ln 的情况

library(lmtest)

library(sandwich)

vcov_hc <- vcovHC(mymodel, type = "HC3")#报错

#针对报错的修正

hat_values <- hatvalues(mymodel)

high_leverage_points <- which(hat_values > 0.2) # 0.2 作为一个常见的阈值

cleaned_data <- trainData[-high_leverage_points, ]

mymodel_cleaned <- lm(happiness ~ religion + study + highest_record + family_income + politics +
                      house_square + health + mood + social + satisfaction_work +
                      house_property + car + marriage + environment + extreme_weather, data =
cleaned_data)

vcov_hc_cleaned <- vcovHC(mymodel_ln_cleaned, type = "HC3")

coefest(mymodel_ln_cleaned, vcov = vcov_hc_cleaned)
```

```
# 加载 ggplot2 包

library(ggplot2)

# 提取残差和预测值

residuals <- residuals(mymodel)

fitted_values <- fitted(mymodel)

# 准备数据框

residual_data <- data.frame(Fitted = fitted_values, Residual = residuals)

# 绘制残差图

ggplot(residual_data, aes(x = Fitted, y = Residual)) +

  geom_point() + # 添加散点

  geom_hline(yintercept = 0, color = "red") + # 添加水平参考线

  labs(title = "Residual Plot", x = "Fitted Values", y = "Residuals") +

  theme_minimal() # 使用简洁的主题

#模型存在异方差问题


library(zoo)

library(lmtest)

bptest(mymodel)


#修正异方差的问题

library(sandwich)

library(lmtest)

# 使用 vcovHC()来获得稳健标准误,此时报错

robust_se <- vcovHC(mymodel, type = "HC1")

hat_vals <- hatvalues(mymodel)

#检验报错值

print(hat_vals[228]) # 查看第 228 个观测值的杠杆值

print(hat_vals[229]) # 查看第 229 个观测值的杠杆值

# 使用 coeftest()展示带有稳健标准误的回归系数

coeftest(model, robust_se)
```